Shansan Gong

hisansas@gmail.com · (+86) 131-6207-6936 · WebPage: https://summmeer.github.io/

Education

Shanghai Jiao Tong University, Bachelor, IE, GPA: 3.84/4.3 Top 5% 2015.09 – 2019.06

Shanghai Jiao Tong University, Master, CS, GPA: 3.78/4 Top 10% 2019.09 – 2022.03

Research Experience

My research interests include Text Generation, Pretrained Language Models and Information Retrieval.

△ Scaling Demonstrations for In-Context Learning 2022.11 – Present

Due to the under-explored long-range language models, current works cannot investigate the in-context performance when we extend the examples for demonstration.

Second author. In-Context Learning with Many Demonstration Examples (ICML 2023 Under Review)

- We propose a pre-trained long-range language model and tune it with instruction. We implement incremental encoding and circular position embedding to ensure the extrapolation and efficiency of the model.
- Experiments show that more demonstration examples bring higher performance for in-context learning.

△ Diffusion Language Model 2022.06 – Present

First author. Diffuseq: Sequence to Sequence Text Generation with Diffusion Models (ICLR 2023)

- Propose a diffusion model DiffuSeq, designed for Seq2Seq text generation tasks.
- Upon extensive evaluation over a wide range of Seq2Seq tasks, we find DiffuSeq achieving comparable or even better performance than six established baselines, including a state-of-the-art model that is based on pre-trained language models. The code is released at here, receiving more than 300 stars.

△ Session-based News Recommendation 2020.06 – 2021.09

News recommendation for anonymous users is challenging because both the lifespan of news articles and the duration of user visits are short. Under the session-based scenario, we model the news recommendation task as the Next-Item Prediction task.

First author. Modeling Implicit Feedback in Session-based News Recommendation (SIGIR 2022)

- We leverage the positive/negative and neutral implicit feedback of the user to figure out to what extend the user likes/dislikes the article, which better tackles the user cold-start problem: we can make an accurate prediction about anonymous users earlier in the session.
- By proper representation of the content and temporal information in the sessions, our approach better handles the article cold-start problem and mitigates the dilemma between diversity and accuracy.

Internship

Microsoft STCA NLP Engineer 2021.06 – 2021.09

In order to optimize the search result, it is necessary to judge whether the intents of different search Query1 and Query2 are consistent, which is formulated as a binary classification task.

- Based on the cross-language pre-training model InfoXLM, introduce PostWeb information (ie, text information such as the title, link, and abstract of the Query search result) as an additional feature of the Query.

♡ Honors and Awards

The Honor of Shanghai Outstanding Graduates Mar. 2022

Shanghai Jiao Tong University Wish Scholarship 2021

2^{nd} *Prize*, China Post-Graduate Mathematical Contest in Modeling Nov. 2020

Shanghai Jiao Tong University Shenzhen Stock Exchange Scholarship 2020